Forens Psychiatr Psychol Kriminol (2023) 17:174–188
https://doi.org/10.1007/s11757-023-00765-9

REVIEW

Structural differences in psychopathy between women and men: a latent modeling perspective

Sandy Sue Spormann[1] · Andreas Mokros[1] · Sabrina Schneider[1]

Received: 1 December 2022 / Accepted: 24 February 2023 / Published online: 18 April 2023
© The Author(s) 2023, corrected publication 2023

Abstract

Research on sex differences in psychopathy indicates that men generally exhibit higher psychopathy scores than women. Measurement equivalence is an important prerequisite for the investigation of mean differences, but is often neglected for psychopathy instruments. The current research provides a systematic qualitative review of the pertinent literature on measurement invariance between men and women for several rater-based and self-report-based psychopathy assessments. Based on 28 studies, we found that the factor structure and factor loadings are most likely comparable between sexes for four out of nine instruments. Results on item thresholds, however, are inconsistent, which questions the comparability of mean scores between men and women for these instruments. The majority of studies that reported acceptable measurement equivalence indicated higher psychopathy scores among men compared to women. As the current literature is neither consistent nor exhaustive, further research needs to address structural differences in psychopathy between biological sexes more systematically.

Keywords Psychopathy · Factor structure · Sex differences · Female psychopathy · Differential item functioning · Assessment

Strukturelle Psychopathie-Unterschiede zwischen Frauen und Männern aus der Sicht latenter Modellierung

Zusammenfassung

Forschungsergebnisse zu Geschlechtsunterschieden im Hinblick auf Psychopathie deuten darauf hin, dass Männer im Allgemeinen höhere Psychopathiewerte aufweisen als Frauen. Messinvarianz der verwendeten Instrumente, die eine wichtige Voraussetzung für die Untersuchung von Mittelwertunterschieden ist, wurde jedoch häufig vernachlässigt. Ziel dieses Artikels ist es, eine systematische Übersicht der einschlägigen Literatur zur Messäquivalenz verschiedener aktenbasierter Instrumente und Selbstberichtsfragebögen zur Erfassung von Psychopathie zu geben. Es wurden 28 Studien in das Review einbezogen. Die Ergebnisse weisen darauf hin, dass die Faktorstruktur und die Faktorladungen bei vier von neun Instrumenten zwischen den Geschlechtern vergleichbar sind. Ergebnisse zu den Item-Schwellenwerten sind jedoch uneinheitlich, was die Vergleichbarkeit der Mittelwerte zwischen Männern und Frauen bei diesen Instrumenten in Frage stellt. Ein Vergleich der Mittelwerte anhand von Studien, in denen Messäquivalenz festgestellt wurde, ergab höhere durchschnittliche Psychopathiewerte bei Männern. Da die Literatur weder konsistent noch erschöpfend ist, sollten zukünftig die strukturellen Unterschiede in Psychopathie zwischen Männern und Frauen systematischer untersucht werden.

Schlüsselwörter Psychopathie · Faktorstruktur · Geschlechtsunterschiede · Weibliche Psychopathie · Differenzielle Itemfunktionen · Psychodiagnostik

In the reference list, articles included in the qualitative review are highlighted with an asterisk.

✉ Sandy Sue Spormann, M.Sc.
sandy.spormann@fernuni-hagen.de

[1] Faculty of Psychology, FernUniversität in Hagen (University of Hagen), Universitaetsstr. 27, 58097 Hagen, Germany

Psychopathic individuals are characterized by a lack of remorse or shame, empathy, and responsibility, but also by high impulsiveness, deceptiveness, poor behavioral control, egocentrism, and the susceptibility to antisocial behavior (Cleckley 1941, 1976). As women commit fewer crimes than men (Leuschner 2020) and are naturally thought to be compassionate, emotional, and selfless (e.g., Connell and Pearse 2015), it is not surprising that, historically, most research on psychopathy focused on males, whereas females have long been neglected in this context. Only in the last two decades has research on female psychopathy increased. Thereby, researchers focused on prevalence rates, external correlates, and manifestations. Lower prevalence rates have almost consistently been found in female compared to male offenders (for a review see Beryl et al. 2014). Typically, among female offenders, base rates of a categorical psychopathy diagnosis (i.e., psychopathic vs. non-psychopathic) are estimated at between 11 and 17% (see Verona and Vitale 2018), whereas in male offenders they are estimated to be about twice as high (i.e., 15–30%; e.g., Hare 2003; Nicholls et al. 2005). Although overall lower prevalence rates are typically found among the general population (Hare 2003), similar disparities are found here, with estimates of, for example, 0.9% for women and 3.7% for men in the UK (Coid et al. 2009). When treated as a dimensional construct, men generally exhibit higher psychopathy scores than women in both institutionalized and non-institutionalized samples (see Verona and Vitale 2018).

In the clinical and criminal justice system, the assessment of psychopathy can have a serious impact on an individual's life and society, as psychopathy is predictive of violence, treatment response, alcoholism, and recidivism (e.g., Douglas et al. 2018; Ellingson et al. 2018; Hare 1999). Therefore, an important concern is whether the observed prevalence differences between men and women rely on actual sex differences in psychopathy, or whether they might—at least in part—be due to sex-related biases in the assessment of those traits. Therefore, it is critical to assure that measurement instruments capture psychopathy equally in men and women, a condition referred to as *measurement equivalence* or *invariance*.

Assessment of psychopathy

Based on Cleckley's (1941) descriptions of the psychopathic individual, Hare (1980) developed the Psychopathy Checklist (PCL) for the clinical and forensic assessment of psychopathy that is currently applied in its revised version (PCL-R; Hare 2003). The PCL-R is referred to as the *gold standard* of the psychopathy assessment, in particular in the clinical and criminal justice system. Both the PCL-R and its screening version, the PCL:SV (Hare et al. 1995), involve a semistructured interview and the review of file information. While the PCL-R is primarily used in the forensic context, the instruments of the Comprehensive Assessment of Psychopathic Personality (CAPP; Cooke et al. 2012) were developed for a variety of settings (e.g., correctional, forensic psychiatric, civil psychiatric, community, and family). Similar to the PCL-R, the CAPP Symptom Rating Scale (CAPP-SRS; Cooke et al. 2012) is an expert rating including, among others, clinical reports, interviews, and behavioral observations. In addition, for research purposes there is the CAPP Lexical Rating Scale (CAPP-LRS; Cooke et al. 2012), which exists in three variants (i.e., the prototypicality, informant, and self-rating forms) and can be used by experts as well as lay people. For both the CAPP and the PCL-R, there are also self-report versions available (CAPP-SR; Sellbom and Cooke 2020; Self-Report Psychopathy Scale in its various editions [SRP]; e.g., Paulhus et al. 2017).

Although expert ratings offer a number of advantages over self-reports, the latter yield useful information regarding the lack of emotional responsiveness in psychopathic individuals, they are economic and easily administered, and reveal response styles.

Beyond the aforementioned self-report instruments (i.e., CAPP-LRS, CAPP-SR, and SRP), other widely used self-reports that assess psychopathy exclusively (Sellbom et al. 2018) include the Psychopathic Personality Inventory(–Revised) (PPI[-R]; Lilienfeld and Widows 2005), the Levenson Self-Report Psychopathy Scale (LSRP; Levenson et al. 1995), the Triarchic Psychopathy Measure (TriPM; Patrick 2010), and the Elemental Psychopathy Assessment (Lynam et al. 2011). Detailed information on the instruments are provided in Table 2 in the Appendix.

Measurement process and measurement invariance

Psychopathy is a psychological construct that cannot be observed directly (and is, therefore, considered a *latent* trait); its behavioral manifestations, however, can be captured by a certain set of items. In Confirmatory Factor Analysis (CFA), the items that make up the latent construct load on a latent factor representing the construct. These factor loadings reflect the strength of the association between the test item and its assigned factor. For example, the four-factor model of psychopathy (Hare and Neuman 2005) was derived from factor analysis of the 20 PCL-R items, whereby each PCL-R item is assigned to a latent factor (i.e., *Interpersonal, Affective, Antisocial*, and *Lifestyle*).

The equivalence of the psychopathy construct across sexes has been validated by several researchers in terms of internal consistency, external correlates, and factor structure

Table 1 Number of studies supporting or rejecting the different levels of measurement invariance

	N	Configural		Metric			Scalar			DIF
		Reject	Full	Reject	Partial	Full	Reject	Partial	Full	
PCL-R	6	1	5	–	2	–	–	2	–	1
PCL:SV	5	2	2	–	–	1	–	–	1[a]	1
CAPP-SRS	0	–	–	–	–	–	–	–	–	–
CAPP-LRS	2	–	2	–	1	1	1	–	1	–
CAPP-SR	0	–	–	–	–	–	–	–	–	–
LSRP	6	–	4	–	1	1	–	–	–	2
SRP (-E, -SF)	5	–	3	1[a]	–	2[a]	–	–	4[a]	–
PPI (-R, -SF)	3	1	1	–	1	–	–	–	–	1
TriPM	2	–	–	–	–	1[a]	–	–	1[a]	1
EPA	0	–	–	–	–	–	–	–	–	–

N number of studies, *PCL–R* Psychopathy Checklist–Revised (Hare 2003), *PCL:SV* Psychopathy Checklist: Screening Version (Hare et al. 1995), *CAPP-SRS* Comprehensive Assessment of Psychopathic Personality–Symptom Rating Scale (Cooke et al. 2012), *CAPP-LRS* Comprehensive Assessment of Psychopathic Personality–Lexical Rating Scale (Cooke et al. 2012), *CAPP-SR* Comprehensive Assessment of Psychopathic Personality–Self-Report (Sellbom and Cooke 2020), *LSRP* Levenson Self-Report Psychopathy Scale (Levenson et al. 1995), *Hare SRP* Hare Self-Report Psychopathy Scale, now formally labeled SRP-4 (Paulhus et al. 2017), *SRP-E* Self-Report Psychopathy Scale–Experimental Version (Williams et al. 2007), *SRP-SF* Self-Report Psychopathy Scale–Short Form (Paulhus et al. 2017), *TriPM* Triarchic Psychopathy Measure (Patrick 2010), *PPI-R* Psychopathic Personality Inventory–Revised (Lilienfeld and Widows 2005), *PPI-SF* Psychopathic Personality Inventory–Short Form (Lilienfeld 1990), *EPA* Elemental Psychopathy Assessment (Lynam et al. 2011), *DIF* differential item functioning
[a]Omitted tests at lower invariance levels

equivalence (for an overview see Verona and Vitale 2018). The latter is usually performed by applying Multi-Group Confirmatory Factor Analysis (MGCFA; e.g., Vandenberg and Lance 2000). If groups exhibit the same number of underlying factors as well as the same factor-item assignment, the measurement is said to be *configurally invariant* (CI). *Metric invariance* (MI) is established if, additionally, the item factor loadings are equal across groups. MI allows comparisons of relations between variables, since the same measurement unit can be assumed. To meaningfully compare mean scores between groups (such as men and women), at least (partial) *scalar invariance* (SI) needs to be established, which means that the thresholds of items are equal across groups. *Strict invariance* can be established by finally constraining residual variances of the items to be equal (Meredith 1993). These steps should be executed by imposing increasingly strict constraints (i.e., from configural to strict; for a review see Vandenberg and Lance 2000). If measurement invariance on either level cannot be obtained, releasing constraints on individual parameters while testing for invariance of the remaining parameters allows to establish partial invariance (Byrne et al. 1989).

The invariance levels can be briefly illustrated on the basis of the *impulsivity* item of the PCL-R: If CI holds, the impulsivity item can be assigned to the same latent factor (i.e., the Lifestyle facet) for both genders. If impulsivity, however, would be assigned to another facet for men than for women (e.g., the Affective facet for men), CI would not hold. MI holds, if impulsivity is an equally adequate indicator of the Lifestyle facet in men and women. If impulsivity, however, was a more direct manifestation of the Lifestyle facet in men than in women (i.e., the factor loading is higher for men), this would suggest a lack of MI. SI holds if mean differences in the impulsivity item between men and women are equal to mean differences in the Lifestyle facet. If, for example, men would be generally more impulsive than women for other reasons than differences in the Lifestyle facet, this would result in a lack of SI. Strict invariance holds if impulsivity assesses the Lifestyle facet with the same precision. If, for example, other factors have a stronger influence on impulsivity in men than in women, this would mean that impulsivity assesses the Lifestyle facet with less precision in men than in women and strict invariance would not hold.

Applying Item Response Theory (IRT; Embretson and Reise 2000; Reise et al. 2005) also allows conclusions to be drawn on the comparability of an instrument across groups. When the probability of endorsing an item differs between groups, the item exhibits so-called *differential item functioning* (DIF). Items that display substantial DIF are of questionable validity and may lead to bias in total scores. Thus, DIF can imply a lack of measurement equivalence.

Current study

The aim of our research was to systematically review the extant literature on measurement invariance of psychopathy instruments between biological sexes. In their review Verona and Vitale (2018) conclude that factor-analytic research with females had produced results largely consistent with studies in males regarding the underlying factor

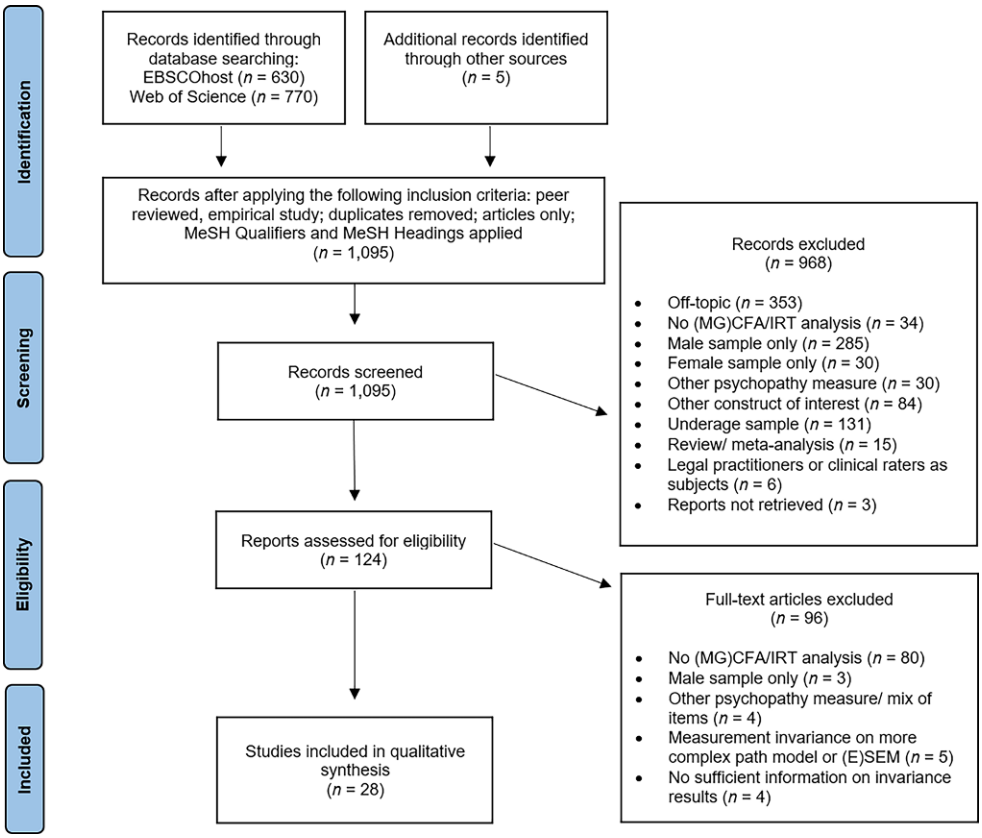

Fig. 1 Flowchart (PRISMA diagram; adapted from Page et al. 2021) of systematic literature. *n* number of reports identified, *MeSH* Medical Subject Headings, *(MG)CFA* (multigroup) confirmatory factor analysis, *IRT* Item Response Theory, *(E)SEM* (exploratory) structural equation modeling

structure of several psychopathy instruments, implying CI. Concerning higher levels of invariance, some of the extant literature suggests SI for several questionnaires (e.g., Neal and Sellbom 2012; Salekin et al. 2014), whereas other research indicates that it only partially holds (e.g., Anestis et al. 2011).

Based on these initial results, we hypothesized that CI holds for all the psychopathy instruments described above (H1). However, prior research led to the expectation that the instruments exhibit different levels of invariance beyond CI (i.e., metric, scalar, and strict invariance) or display DIF (H2). In addition, it was assumed that latent factor mean scores are generally higher in men than in women (H3)[1].

Methods

This review applied a systematic qualitative approach with a selection process following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al. 2021). The literature research was conducted in the following EBSCOhost sources in September 2022: APA PsycArticles, APA PsycInfo, PSYNDEX Literature with PSYNDEX Tests, and Psychology and Behavioral Sciences Collection. Moreover, the search was repeated in the Web of Science (Social Science Citation Index [SSCI]) database. The search string comprised the names of the above-mentioned instruments—which were selected mainly because they exclusively measure psychopathy—as well as terms related to *measurement invariance* and *sex*.

Documentation of the whole search (including a table with all screened reports) and the supplemental material can be found on the Open Science Framework (OSF): https://osf.io/dk2q7/. All steps of the review, including the hypotheses, were preregistered on the OSF.

Only empirical studies published in peer-reviewed journals were considered. Book chapters, dissertation abstracts, meta-analyses, and reviews were omitted from the search. Only studies that included both males and females were considered to allow for direct comparisons. Hereby, we also included studies that estimated CFAs separately for men and women (instead of MGCFA) since adequate model fit in both groups can be indicative of CI. A further inclusion criterion was the participants' age (minimum 18 years), since personality disorders are not diagnosed in underage individuals (e.g., American Psychiatric Association 2013). As an exception, we included two articles (Adams et al.

[1] Note that differences in latent means were only considered if the respective researchers reported (partial) SI for a given psychopathy measure. If full SI is obtained, observed mean scores and latent mean scores are thought to be sufficiently equal to interpret the observed mean scores meaningfully. That was the case for two studies, for which we report observed mean scores (Neumann and Hare 2008; Walsh et al. 2019).

2020; Gummelt et al. 2012) with samples that mainly consisted of adults but also 17-year-old individuals.

The search yielded 630 results in EBSOhost and 770 in the Web of Science. Five articles were detected from backward search in the included articles. Duplicates were removed, leading to a total of 1095 articles, most of which were excluded upon screening. The most frequent reasons for exclusion were clearly off-topic articles ($n = 353$) and samples including only males ($n = 285$) or underage individuals ($n = 131$). Further reasons for exclusion are documented in the PRISMA diagram (Fig. 1). The final literature review comprised 28 articles, which are highlighted in the "Reference" section. Detailed information on the samples, methods, and main results of those studies can be obtained from Table 3 in the Appendix.

Results

This section summarizes the main findings of our qualitative synthesis, grouped by psychopathy measure. None of the studies included in this review investigated the latent structure of the EPA, the CAPP-SRS, or the CAPP-SR. Moreover, no study assessed strict invariance. An overview of the number of studies supporting invariance on the respective levels is given in Table 1. Some researchers omitted tests for invariance on one or more levels or did not report according results; therefore, the number of studies varies depending on the invariance level.

PCL-R

Six studies examined measurement equivalence of the PCL-R. Four of them estimated CFAs separately for males and females, two studies applied MGCFA, and one study IRT analyses (cf. Table 3 in the Appendix).

Five of the six pertinent studies found support for CI (Bolt et al. 2004; Klein Haneveld et al. 2022; Neumann et al. 2007; Walters et al. 2011; Windle and Dumenci 1999), whereas one did not find empirical support for their model in either subsample (Darke et al. 1998).

Three studies examined higher levels of invariance. Both Klein Haneveld et al. (2022) and Windle and Dumenci (1999) found partial MI with one PCL-R item lacking invariance. Their MGCFA results further supported partial SI, with three non-invariant item thresholds reported in both studies. In their IRT study, Bolt et al. (2004) found DIF for 12 items. Since the magnitude of the detected item differences between the male and female offender groups seemed to be negligible, the authors concluded that partial SI was confirmed for the PCL-R in their samples.

With regard to mean differences, females scored significantly lower than males on all four PCL-R facets in the study by Klein Haneveld et al. (2022; *interpersonal,* estimate[2] = –0.549, $p = 0.002$; *affective,* estimate = –0.689, $p < 0.001$; *lifestyle,* estimate = –0.612, $p < 0.001$; *antisocial,* estimate = –1.498, $p < 0.001$). Likewise, the overall latent mean score was higher for men in Bolt et al. (2004, estimate = –0.45, $p < 0.001$). In contrast, Windle and Dumenci (1999) did not find latent mean differences between men and women on either factor in alcoholic inpatients ($p > 0.10$).

In summary, results of the studies reviewed largely support lower (i.e., CI) and higher levels (i.e., partial SI) of measurement invariance for the PCL-R. However, results on sex differences in the psychopathic traits assessed were inconsistent.

PCL:SV

Five studies investigated the latent structure of the PCL:SV, whereby one study applied Exploratory Factor Analysis (EFA), two studies separate CFA in men and women, two studies MGCFA, and one IRT analysis (cf. Table 3 in the Appendix). Two studies found evidence of CI (Skeem et al. 2003; Thomson et al. 2019), whereas one did not (Forth et al. 1996). An EFA by Strand and Belfrage (2005) revealed a two-factor structure of the PCL:SV for males and a three-factor structure for females.

Only one of the three studies examining higher levels of invariance tested and found support for full MI (Skeem et al. 2003). Skeem et al. (2003) did not further test for SI. Results by Neumann and Hare (2008) supported full SI, while in the IRT study by Strand and Belfrage (2005) five PCL:SV items displayed DIF. Neumann and Hare (2008) were the only researchers to report mean differences, and they found higher observed PCL:SV scores for men ($M = 3.53$, $SD = 3.79$) than for women ($M = 2.16$, $SD = 3.23$) with a small effect size ($d = 0.30$).

In sum, results do not unanimously support the presence of CI. Studies that tested for higher invariance levels (i.e., MI and SI) attest to the measurement equivalence of PCL:SV items in men and women. Yet, some PCL:SV items may still show sex-related response bias according to IRT analysis. The existence of sex differences in PCL:SV-assessed psychopathy traits (H3) is supported, but evidence is limited to a single study.

CAPP-LRS

Two studies included in this review investigated the CAPP-LRS self-rating form with MGCFA (Hanniball et al. 2021; Sellbom et al. 2015) and both found support for CI. One

[2] Latent mean estimates for women compared to the male sample in which latent means were fixed to 0.

study established full MI (Sellbom et al. 2015) and one partial MI, with nine non-invariant item loadings (Hanniball et al. 2021). Only Sellbom et al. (2015) were able to additionally establish SI, while Hanniball et al. (2021) concluded that SI was not present. In the study by Sellbom et al. (2015), men scored significantly higher than women on all factors with small to medium effect sizes (general factor, $z = 8.14$, $p < 0.001$, $d = 0.68$, bi-factors *boldness/emotional stability,* $z = 2.84$, $p = 0.005$, $d = 0.28$; *emotional detachment,* $z = 4.86$, $p < 0.001$, $d = 0.60$; and *disinhibition,* $z = 1.97$, $p = 0.049$, $d = 0.18$).

Taken together, extant findings indicate lower levels of measurement invariance for the self-report version of the CAPP-LRS. However, as results on SI are inconclusive, the reported sex differences in CAPP-LRS scores (Sellbom et al. 2015) should be interpreted with caution.

LSRP

In total, six studies examined measurement equivalence for the LSRP. There was one study providing CFA results and one providing congruence coefficients—an index of the similarity between factors. Two studies applied MGCFA and two IRT, respectively (cf. Table 3 in the Appendix). CI was supported by all four studies that examined CI (Anestis et al. 2019; Lynam et al. 1999; Sellbom 2011; Somma et al. 2014).

Of the two studies that tested for MI, one found support for full MI (Sellbom 2011) and one for partial MI (Lynam et al. 1999) with one non-invariant factor loading. Both studies did not further test for SI. The two studies that applied IRT detected different degrees of DIF: In Gummelt et al. (2012), 17 items displayed DIF between men and women, while in Hauck-Filho and Teixeira (2014), only 3 items displayed DIF.

All in all, the studies reviewed support the notion of measurement equivalence for the LSRP at the first two invariance levels. Nevertheless, the results of the DIF analyses indicate that the number of items that work differently in men and women may be sample dependent. (Latent) mean differences in LSRP scores between men and women were not reported.

SRP

Five studies examined the Hare SRP (Neal and Sellbom 2012), the experimental SRP (SRP-E; Neumann et al. 2012), or the SRP-Short Form (SRP-SF; Carre et al. 2018; Dotterer et al. 2017; Walsh et al. 2019) with MGCFA. CI was examined and supported by three of these studies (Dotterer et al. 2017; Neumann et al. 2012; Walsh et al. 2019). Three studies tested for MI, two of which reported for the SRP to attain full MI (Neal and Sellbom 2012; Neumann et al. 2012), whereas Carre et al. (2018) could not. Four studies attained SI (Dotterer et al. 2017; Neal and Sellbom 2012; Neumann et al. 2012; Walsh et al. 2019). Men ($M = 62.10$, $SD = 15.95$) obtained higher observed SRP-SF total scores than women ($M = 51.97$, $SD = 15.19$; $F(1, 587) = 65.85$, $p < 0.001$) with a small effect size ($\eta^2 = 0.10$; Walsh et al. 2019). Taken together, the majority of studies found evidence of measurement equivalence for the SRP in men and women, at both lower and higher invariance levels. Our prediction on sex differences in SRP scores was also supported, but only by a single study.

PPI

Of the three studies examining equivalence for the PPI, two studies included separate CFAs for men and women, one of which used additional MGCFA, and one used IRT (cf. Table 3 in the Appendix). Anestis et al. (2011) compared three competing models of the PPI-R and found that both the one-factor (PPI-*Psychopathy*) and the two-factor model (*Self-Centered Impulsivity* and *Fearless Dominance*) yielded good fit for the female sample but barely acceptable fit for the male sample, whereas for the three-factor model, fit was modest for both groups. In contrast, Adams et al. (2020) did not find empirical support for their eight-factor PPI-SF model, neither in male, nor in female participants.

Two studies tested for higher levels of invariance. Anestis et al. (2011) reported partial MI for the one- and the two-factor PPI model with two items and one non-invariant item, respectively. The partially constrained three-factor model fit the data poorly. Anestis et al. (2011) did not apply further constraints. In the IRT study, 61.1% ($n = 80$) of the PPI-R items displayed DIF across sex groups (Eichenbaum et al. 2019).

The results cast doubt on the presence of measurement invariance for the PPI and the general suitability of the respective measurement models tested. Moreover, the amount of DIF was substantial. Mean sex differences in PPI scores were not reported in any of the studies reviewed.

TriPM

One study investigated the TriPM by means of MGCFA and one by IRT. Neither study tested for CI. Full MI was attained by Carre et al. (2018). The same study further supported full SI. In the study by Eichenbaum et al. (2021), 61% of the TriPM items ($n = 34$) displayed DIF. Women scored lower than men on all three factors in the study by Carre et al. (2018; *boldness,* $t(474) = 5.874$, $p < 0.001$; *meanness,* $t(474) = 8.262$, $p < 0.001$, and *disinhibition,* $t(474) = 3.898$, $p < 0.001$). Although the MGCFA results reported support the presence of higher-level mea-

surement invariance for the TriPM, IRT results imply a lack of equivalence. Thus, evidence of sex differences in TriPM scores need to be interpreted with caution, not least because they have only been reported by one study.

Discussion

Within this qualitative synthesis we examined equivalence across sexes in the latent structure of psychopathy as measured with several expert and self-report assessment instruments. We hypothesized that the instruments exhibit at least a basic level of measurement invariance (i.e., CI [H1]). In line with previous research (Verona and Vitale 2018), CI was confirmed for several instruments, that is the PCL-R, the CAPP-LRS, the LSRP, and the SRP. No study that investigated the TriPM reported results on CI. Results on the PCL:SV and the PPI-R were inconsistent, but, in sum, suggest a lack of CI. A general reexamination of the latent factor structure of the PPI in male and female samples should be considered. Assessments of measurement equivalence of the EPA, the CAPP-SRS, and the CAPP-SR are still lacking. Therefore, we found conclusive evidence of configural measurement invariance for four of the psychopathy measures addressed (i.e., PCL-R, CAPP-LRS, LSRP, SRP), which provides only partial empirical support for our first hypothesis.

We further assumed that the psychopathy instruments would exhibit different levels of measurement invariance or display DIF (H2). Based on the studies included, (partial) MI was mostly confirmed for all instruments except the EPA, CAPP-SRS, and CAPP-SR. With regard to the SRP, however, results on MI were mixed. With respect to the remaining instruments results were limited to one or two studies each. In view of this sparse empirical evidence, the present results should be interpreted with caution.

Findings on scalar invariance (SI) are highly inconclusive between but also within several instruments. Even though Bolt et al. (2004) found 12 items of the PCL-R to display DIF, partial SI was largely supported. Likewise, studies unanimously support the presence of SI for the SRP measures. No clear conclusions can be drawn for the other psychopathy scales reviewed, as the results on SI for those measures were inconsistent and partly appeared to depend on the statistical method (e.g., DIF analysis vs. MGCFA). In sum, six instruments appeared to exhibit (partial) MI, whereas SI can only be presumed for two measures, the PCL-R and the SRP, with relative certainty.

Finally, we assumed that men exhibited higher levels of psychopathy (H3) when the assessment method has been proven to be comparable for men and women, i.e., when at least partial SI has been confirmed for that measure. Across different samples (i.e., community, student, incarcerated) six studies (Bolt et al. 2004; Carre et al. 2018; Klein Haneveld et al. 2022; Neumann and Hare 2008; Sellbom et al. 2015; Walsh et al. 2019) found either latent or observed mean scores of the respective instrument to be higher for men than for women with small to medium effect sizes (please note that effect sizes were only given for three studies). In contrast, only one study (Windle and Dumenci 1999) did not find any significant sex differences; however, this might be attributed to the specific nature of the sample tested, which comprised alcoholic inpatients. In sum, these results support our last hypothesis and corroborate previous findings (Verona and Vitale 2018).

It is conceivable that the inconsistent results with regard to both invariance and mean differences are, at least in part, due to sampling issues. Some samples comprised criminal offenders (e.g., Bolt et al. 2004; Sellbom 2011), methadone patients (Darke et al. 1998), or alcoholics (Windle and Dumenci 1999), whereas others comprised community individuals (Neumann and Hare 2008; Somma et al. 2014) or undergraduate students (e.g., Gummelt et al. 2012; Lynam et al. 1999). The contradictory results on the PCL:SV, for example, may be attributed to differences in sample sizes between the studies by Forth et al. (1996, $n = 75$ per sex) and Thomson et al. (2019; $N = 565$). Likewise, sampling issues were apparent for the CAPP-LRS, for which qualitatively different samples were compared (general population [Sellbom et al. 2015] vs. felons [Hanniball et al. 2021]).

Another sample property that might cause inconclusive results is ethnicity or cultural background. The majority of the studies has been conducted in Western samples, whereas Hauck-Filho and Teixeira (2014) examined sex differences for the LSRP in a Latin American sample and obtained results that differed from those of a U.S. sample with the same method (Gummelt et al. 2012). Hence, the psychopathy measures examined were not only administered in different cultural contexts, but also sometimes in different languages, which has a known impact on measurement invariance analyses (see Bader et al. 2021). Accordingly, upon comparing women from different world regions in their large worldwide sample, Neumann et al. (2012) found that SI of a given psychopathy measure depended on the world regions that women came from. In this context, it is important to mention that exclusively German samples were not investigated in any of the 28 studies (but German subsamples are included in Neumann and Hare [2008] and Walters et al. [2011]). In order to draw conclusions about the use of the investigated instruments for both biological sexes in Germany, this research gap needs to be addressed.

Besides sample properties—and sample heterogeneity in particular—as a potential cause of inconsistent findings, there are five methodological aspects that we would like to address. First, it should be taken into account that the various psychopathy instruments have been developed on

the basis of different conceptualizations of psychopathy. Whether or not an instrument is invariant between genders may depend on the underlying theoretical concept. Second, if a confirmatory model fits the data, it does not necessarily mean that another model would not fit the data even better. Thus, it is possible that the optimal measurement model of some instruments differs between men and women (e.g., Salekin et al. 1997). Third, many of the studies we reviewed struggled with the implementation of a properly fitting measurement model. To achieve acceptable fit, the models were often adapted in various ways, which raises the probability of *overfitting*. These discrepancies in model adaptation may have caused differences in the subsequent MI analyses as well. Fourth, a large amount of studies only attained partial MI or partial SI, but neither provided explanations for the partial non-invariance nor compared further results to those of a model set to invariance, as it should be done to investigate the potential effects of partial non-invariance (Putnick and Bornstein 2016).

Finally, a major concern is the inconsistent execution and reporting of the analytic strategy and results across studies. Study authors sometimes omitted tests of one or more invariance levels. They relied on different indices (e.g., likelihood ratio tests of nested models, or changes [Δ] in comparative and absolute model fit indices) and applied inconsistent cut-off values. Among the criteria applied, some were stricter than others, thus, affording ambiguous conclusions about measurement invariance. Meredith (1993) recommended to execute MGCFA with increasing restrictiveness. Moreover, there are clear conventions on testing and reporting invariance (Putnick and Bornstein 2016), which should be applied more consistently in future studies. In order to circumvent the drawbacks of sample size-dependent significance tests and cut-off values, Nye and Drasgow (2011) suggested the application of effect size estimates.

Taken together, the results on sex differences in the covariance and mean structure of psychopathy measures are inconclusive. The initial question whether observed sex differences in psychopathy occur because of actual trait differences or whether they are due to sex-related bias in the assessment method(s) cannot be answered clearly. Nevertheless, the substantial amount of studies attaining partial SI for at least some of the measures provides a promising perspective towards a future assessment of psychopathy that would be equally applicable to both men and women. Future research should consider a stringent investigation of the various instruments, as for example presented by Klein Haneveld et al. (2022). In this regard, the conditions and reasons for (partial) non-invariance need to be addressed directly, for example, by systematically assessing different target populations. Only based on such studies will it be possible to make informed decisions on the suitable measure(s) when exploring psychopathy in men and women. Until then, forensic practice should be cautious when interpreting mean scores and applying cut-off values in women, as different norms may apply to them.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix

Table 2 Number of items, reliability, procedure, and duration of the included psychopathy instruments

Instrument	Items	Reliability (source)	Procedure	Duration
Comprehensive Assessment of Psychopathic Personality–Lexical Rating Scale (Cooke et al. 2012)	33	α = 0.96 (for the self-report; Sellbom et al. 2015)	Self-report, informant report, or prototypicality rating	15–20 min
Comprehensive Assessment of Psychopathic Personality–Self-Report (Sellbom and Cooke 2020)	99	α = 0.95 (Sellbom and Cooke 2020)	Self-report	15–20 min
Comprehensive Assessment of Psychopathic Personality–Symptom Rating Scale (Cooke et al. 2012)	33	ICC = 0.56; α = 0.96 (Pedersen et al. 2010)	Semistructured interview (with assessee or informant) and file review	3–8 h
Elemental Psychopathy Assessment (Lynam et al. 2011)	178	α = 0.95 (Lynam et al. 2011)	Self-report	10–20 min
Elemental Psychopathy Assessment–Short Form (Lynam et al. 2013)	72	α = 0.74 (averaged across scales, Lynam et al. 2013)	Self-report	5–10 min
Levenson Self-Report Psychopathy Scale (Levenson et al. 1995)	26	α = 0.83 (Miller et al. 2008)	Self-report	5 min
Psychopathic Personality Inventory–Revised (Lilienfeld and Widows 2005)	154	α = 0.90 (Lilienfeld and Widows 2005)	Self-report	20–45 min
Psychopathic Personality Inventory–Short Form (Lilienfeld 1990)	56	α = 0.85 (Lilienfeld and Hess 2001)	Self-report	10–15 min
Psychopathy Checklist–Revised (Hare 2003)	20	ICC = 0.87–0.93 (Hare 2003)	Semistructured interview and file review	2–4 h
Psychopathy Checklist–Screening Version (Hare et al. 1995)	12	ICC = 0.80–0.89; α = 0.89 (Pedersen et al. 2010)	Semistructured interview and file review	1.5 h
Self-Report Psychopathy Scale 4th Edition (Paulhus et al. 2017)	64	α = 0.92 (Neal and Sellbom 2012)	Self-report	12 min
Self-Report Psychopathy Scale–Experimental Version (SRP-E)	40	α = 0.88 (Williams et al. 2007)	Self-report	5–10 min
Self-Report Psychopathy Scale–Short Form (Paulhus et al. 2017)	29	α = 0.82 (Walsh et al. 2019)	Self-report	3–10 min
Triarchic Psychopathy Measure (Patrick 2010)	58	α = 0.91 (Yoon et al. 2022)	Self-report	10–20 min

ICC intraclass correlation as measure of interrater reliability (ICC ≥ 0.75 represents excellent, 0.40 ≤ ICC ≤ 0.75 fair to good, and ICC ≤ 0.40 poor reliability [Fleiss 1986]), *α* Cronbach's coefficient of internal consistency regarding self-report measures (α ≥ 0.90 represents excellent, 0.90 > α ≥ 0.80 good, 0.80 > α ≥ 0.70 acceptable, 0.70 > α ≥ 0.60 questionable, 0.60 > α ≥ 0.50 poor, and α < 0.50 unacceptable reliability [George and Mallery 2003])

Table 3 Overview of the included studies with information on samples, methods, and main results

Study	Hypotheses	Sample	Instrument	Method	Measurement model	Main findings
Bolt et al. (2004)	H1, H2, H3	3847 male criminal offenders 1219 female criminal offenders	PCL-R	IRT, CFA	Two-factor, four-facet hierarchical model with an additional third-order factor	The CFA model yielded acceptable fit for both male and female subsamples, indicating CI. DIF was found for 12 items. Latent mean scores were higher for men than for women (estimate = –0.45, $p < 0.001$)
Darke et al. (1998)	H1	200 community and 176 prison methadone program participants (65% male; M_{age} = 32.5 [SD = 6.2, range 19–49])	PCL-R	CFA	Two-factor model	The model fit both male and female subsamples poorly
Neumann et al. (2007)	H1	4865 male (M_{age} = 20.03 [SD = 7.30]) offenders 1099 female offenders (M_{age} = 17.24 [SD = 6.87])	PCL-R	CFA	Four-correlated factors model	The model fit both male and female subsample well, indicating CI. Discrimination (loadings) and extremity (intercept) parameters indicated a sex bias on three or two items, respectively
Walters et al. (2011)	H1	a) 473 male released psychiatric patients (M_{age} = 29.8 [SD = 6.3]) b) 356 female released psychiatric patients (M_{age} = 29.9 [SD = 6.1])	PCL-R	CFA	Four-factor model	Model fit was acceptable for both samples, indicating CI
Klein Haneveld et al. (2022)	H1, H2, H3	110 female and 147 male criminal offenders	PCL-R	MGCFA	Two-factor, four-facet hierarchical model with an additional third-order factor	Full CI was obtained. Partial MI held with one item and partial SI held with two items differing between sexes. Females scored significantly lower than males on all four facets with the largest effect on the antisocial facet and the smallest effect on the interpersonal facet. (*interpersonal,* estimate = –0.549, $p = 0.002$; *affective,* estimate = –0.689, $p < 0.001$; *lifestyle,* estimate = –0.612, $p < 0.001$, *antisocial,* estimate = –1.498, $p < 0.001$)
Windle and Dumenci (1999)	H1, H2, H3	481 male and 321 female alcoholic psychiatric inpatients (M_{age} = 34.41 [SD = 7.71])	PCL-R	MGCFA	Two-factor model	CI fully held, partial MI held with one item differing between sexes, and partial SI held with three items differing between sexes. No latent mean differences on either of the two factors were found ($p > 0.10$)
Strand and Belfrage (2005)	H1, H2	129 female and 499 male Swedish offenders (M_{age} = 36 [SD = 11, range 18–68])	PCL:SV	EFA and IRT	–	EFA revealed a two-factor solution for males and a three-factor solution for females. DIF was found for 5 items
Forth et al. (1996)	H1	1) 25 male and 25 female undergraduate students (M_{age} = 24.8 [SD = 6.7]) 2) 25 male and 25 female undergraduate students (M_{age} = 20.2 [SD = 2.4]) 3) 25 male and 25 female undergraduate students (M_{age} = 21.6 [SD = 5.1])	PCL:SV	Congruence coefficients	Two-factor model	A clear two-factor solution was not obtained in either sample. The congruence coefficient indicated that factors were not satisfyingly similar between sexes (r_c = 0.83 for the first factor and r_c = 0.34 for the second factor)

Table 3 (Continued)

Study	Hypotheses	Sample	Instrument	Method	Measurement model	Main findings
Thomson et al. (2019)	H1	385 male and 180 female adults (M_{age} = 27.5 [SD = 6.15, range 18–45])	PCL:SV	CFA	A two-factor and a four-facet model	Both models exhibited adequate fit to the data for both sexes, indicating CI
Neumann and Hare (2008)	H2, H3	Community sample of 196 males and 318 females (M_{age} = 31.0 [SD = 6.1])	PCL:SV	MGCFA	Four correlated factors model	SI model exhibited good fit to the data. Other levels of invariance were not tested. Observed PCL:SV scores were higher for men (M = 3.53 [SD = 3.79]) than for women (M = 2.16, [SD = 3.23]) with a small effect size (d = 0.30)
Skeem et al. (2003)	H1, H2	870 psychiatric inpatients (58% male, M_{age} = 30 [SD = 6])	PCL:SV	MGCFA	Three-factor model	CI and MI held across sexes. SI was not tested
Hanniball et al. (2021)	H1, H2	1414 individuals endorsing serious criminal histories (62% male, M_{age} = 30.34 [SD = 7.39, range 18–73])	CAPP-LRS self-report	MGCFA	First-order three-factor model	CI held, partial MI held, but SI could not be established
Sellbom et al. (2015)	H1, H2, H3	Community sample of 719 individuals (51% female, age range 18–> 70)	CAPP-LRS self-report	MGCFA	Bi-factor model with one general factor and three bi-factors	CI held, partial MI held with 12 items differing between sexes. Partial SI held with 7 items differing between sexes. Men scored significantly higher than women on the general factor (z = 8.14, p < 0.001, d = 0.68) and the three bi-factors *boldness/emotional stability* (z = 2.84, p = 0.005, d = 0.28), *emotional detachment* (z = 4.86, p < 0.001, d = 0.60), and *disinhibition* (z = 1.97, p = 0.049, d = 0.18)
Anestis et al. (2019)	H1	840 military personnel (81.2% male, M_{age} = 27.08)	LSRP	CFA	26-items three-factor model, 19-items two- and three-factor models	The models all provided acceptable fit for both male and female subsamples, suggesting CI for all models
Somma et al. (2014)	H1	Community sample of 277 males and 463 females (M_{age} = 36.52 [SD = 12.87, range 18–65])	LSRP	Congruence coefficients	–	Congruence coefficients supported the fit of a three-factor model across sex groups (r_c values of 0.97, 0.94, and 0.94 for factors 1, 2, and 3, respectively), indicating CI
Lynam et al. (1999)	H1, H2	Mixed subsample of 1746 undergraduate students	LSRP	MGCFA	Two-factor model	Partial MI with one item loading differing between sexes. No further analyses were reported
Sellbom (2011)	H1, H2	1) 558 male prison inmates (M_{age} = 32.31 [SD = 9.65, range 18–66]) 2) 202 male and 200 female college students (M_{age} = 19.39 [SD = 3.37, range 18–56])	LSRP	CFA, MGCFA	One-factor model, two different two-factor models, two different three-factor models	CFAs did not yield adequate fit for any of the models. Two models were modified and yielded adequate fit afterwards. CI and MI were obtained for the revised three-factor model. SI was not examined
Gummelt et al. (2012)	H2	1516 college students (58.9% female, M_{age} = 18.68 [SD = 1.8, range 17–46])	LSRP	IRT	–	Seventeen items displayed DIF

Table 3 (Continued)

Study	Hypotheses	Sample	Instrument	Method	Measurement model	Main findings
Hauck-Filho and Teixeira (2014)	H2	Brazilian community sample of 644 females and 240 males (M_{age} = 28.58 [SD = 10.55])	LSRP	IRT	–	Three items displayed DIF with negligible differences in difficulty estimates
Neal and Sellbom (2012)	H1, H2	428 female and 178 male undergraduate students (M_{age} = 19.90 [SD = 3.48, range 18–48])	Hare SRP	MGCFA	Four-factor model	Full MI and SI between sexes were supported
Neumann et al. (2012)	H1, H2	19,183 females and 13,833 males from 58 nations (M_{age} = 21 [range 18–91])	SRP-E (with 19 instead of 21 items)	CFA; MGCFA	Four-factor model	CFAs supported good model fit for both sexes. Full CI, MI and SI across sex were supported
Carre et al. (2018)	H1, H2, H3	481 workers recruited from MTurk (61.3% female, M_{age} = 36.10 [SD = 12.58])	SRP-SF, TriPM	MGCFA	SRP: four-factor model TriPM: three-factor model	No MI was obtained for the SRP-SF. MI and SI held for the TriPM. Latent means were not equivalent ($\Delta X^2 = 79.72$, $p < 0.001$). Hereby, women scored lower than men on all three factors (*boldness,* $t(474) = 5.874$, $p < 0.001$; *meanness,* $t(474) = 8.262$, $p < 0.001$, and *disinhibition,* $t(474) = 3.898$, $p < 0.001$)
Dotterer et al. (2017)	H1, H2	1) 384 college students (65.1% female, M_{age} = 19.3 [SD = 1.67, range 18–34]) 2) 848 college students (45.8% female, M_{age} = 20.71 [SD = 4.24, range 18–57]) 3) 1012 college students (57.3% female, M_{age} = 19.67 [SD = 1.25, range 18–22]) 4) 283 22-year-old men from the Pitt Mother & Child Project	SRP-SF	CFA, MGCFA	Four-correlated factor model, bi-factor model with one general factor and four bi-factors	CFAs for the four-correlated factor model yielded inadequate fit for men, but good fit for women. The four bi-factor model yielded good fit for both sexes. CI was not established for the four-correlated factor model but for the bi-factor model. At the scalar level there is evidence for the presence of SI, the X^2-test, however, was significant
Walsh et al. (2019)	H1, H2, H3	590 undergraduate students (65% females, M_{age} females = 20.52 [SD = 3.89], M_{age} males = 21.07 [SD = 4.92])	SRP-SF	MGCFA	Four-factor model	Full CI and full SI were obtained. MI was not examined. Men (M = 62.10, SD = 15.95) obtained higher observed SRP-SF total scores than women (M = 51.97, SD = 15.19; $F(1, 587) = 65.85$, $p < 0.001$) with a small effect size ($\eta^2 = 0.10$)
Eichenbaum et al. (2019)	H2	348 male and 704 female college students (age range 18–49)	PPI-R	IRT	–	Eighty items (61.1%) displayed significant DIF between men and women
Anestis et al. (2011)	H1, H2	145 male and 215 female undergraduates (M_{age} = 18.95 [SD = 1.42, range 18–28])	PPI-R	CFA; MGCFA	One-factor, two-factor, and three-factor model	In separate CFAs the one-factor and two-factor model yielded good fit for the female sample and barely acceptable fit for the male sample, while for the three-factor model fit was modest for both groups. Partial MI was attained for the one- and the two-factor model with two items and one item differing between sexes, respectively. The partially constrained three-factor model fit the data poorly. No further constraints were applied

Table 3 (Continued)

Study	Hypotheses	Sample	Instrument	Method	Measurement model	Main findings
Adams et al. (2020)	H1	Community sample of 869 females and 378 males (M_{age} = 19.32 [SD = 2.31, range 17–51])	PPI-SF	CFA	Eight-factor model	The model fit both male and female subsamples poorly
Eichenbaum et al. (2021)	H2	1) 138 male and 269 female students 2) 272 males and 258 females recruited from MTurk (M_{age} = 30.3 [SD = 12.2, range 18–75])	TriPM	IRT	–	Thirty-four items (61%) displayed significant DIF between men and women

PCL–R Psychopathy Checklist–Revised (Hare 2003), *IRT* item response theory, *CFA* confirmatory factor analysis, *CI* configural invariance, *DIF* differential item functioning, *MGCFA* multigroup confirmatory factor analysis, *MI* metric invariance, *SI* scalar invariance, *PCL:SV* Psychopathy Checklist: Screening Version (Hare et al. 1995), *EFA* exploratory factor analysis, *CAPP-LRS* Comprehensive Assessment of Psychopathic Personality–Lexical Rating Scale (Cooke et al. 2012), *LSRP* Levenson Self-Report Psychopathy Scale (Levenson et al. 1995), *Hare SRP* Hare Self-Report Psychopathy Scale, now formally labeled SRP-4 (Paulhus et al. 2017), *SRP-E* Self-Report Psychopathy Scale–Experimental Version (Williams et al. 2007), *MTurk* Amazon's Mechanical Turk, *SRP-SF* Self-Report Psychopathy Scale–Short Form (Paulhus et al. 2017), *TriPM* Triarchic Psychopathy Measure (Patrick 2010), *PPI-R* Psychopathic Personality Inventory–Revised (Lilienfeld and Widows 2005), *PPI-SF* Psychopathic Personality Inventory–Short Form (Lilienfeld 1990)

Acknowledgements The authors would like to thank Christoph Heine for his helpful advice, review of the manuscript, and help with table editing. We also thank Sabrina Kornberger and Anne-Marie Kutsche for their literature research on the properties of the instruments and the peer-reviewers for their helpful comments on the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Conflict of interest A. Mokros receives royalties from the sale of the German version of the PCL-R. S.S. Spormann and S. Schneider declare that they have no competing interests.

References

Adams EW, Bontemps AP, Decker K et al (2020) Model fit and convergent and discriminant validity of the psychopathic personality inventory-short form: A comparison of competing models. J Psychopathol Behav Assess 42:647–665. https://doi.org/10.1007/s10862-020-09829-0 (*)

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders: DSM-5, 5th edn. American Psychiatric, Washington https://doi.org/10.1176/appi.books.9780890425596

Anestis JC, Caron KM, Carbonell JL (2011) Examining the impact of gender on the factor structure of the Psychopathic Personality Inventory—Revised. Assessment 18:340–349. https://doi.org/10.1177/1073191111403243 (*)

Anestis JC, Green BA, Arnau RC, Anestis MD (2019) Psychopathic personality traits in the military: An examination of the Levenson Self-Report Psychopathy Scales in a novel sample. Assessment 26:670–683. https://doi.org/10.1177/1073191117719511 (*)

Bader M, Jobst LJ, Zettler I et al (2021) Disentangling the effects of culture and language on measurement noninvariance in cross-cultural research: The culture, comprehension, and translation bias (CCT) procedure. Psychol Assess 33:375–384. https://doi.org/10.1037/pas0000989

Beryl R, Chou S, Völlm B (2014) A systematic review of psychopathy in women within secure settings. Pers Individ Dif 71:185–195. https://doi.org/10.1016/j.paid.2014.07.033

Bolt DM, Hare RD, Vitale JE, Newman JP (2004) A multigroup item response theory analysis of the Psychopathy Checklist-Revised. Psychol Assess 16:155–168. https://doi.org/10.1037/1040-3590.16.2.155 (*)

Byrne BM, Shavelson RJ, Muthén B (1989) Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychol Bull 105:456–466. https://doi.org/10.1037/0033-909.105.3.456

Carre JR, Mueller SM, Schleicher KM, Jones DN (2018) Psychopathy and deviant workplace behavior: A comparison of two psychopathy models. J Personal Disord 32:242–261. https://doi.org/10.1521/pedi_2017_31_296 (*)

Cleckley H (1941) The mask of sanity. Mosby, Maryland Heights

Cleckley H (1976) The mask of sanity, 5th edn. Mosby, Maryland Heights

Coid JW, Yang M, Ullrich S et al (2009) Prevalence and correlates of psychopathic traits in the household population of Great Britain. Int J Law Psychiatry 32:65–73. https://doi.org/10.1016/j.ijlp.2009.01.002

Connell R, Pearse R (2015) Gender: In world perspective, 3rd edn. Polity Press, Cambridge

Cooke DJ, Hart SD, Logan C, Michie C (2012) Explicating the construct of psychopathy: Development and validation of a conceptual model, the Comprehensive Assessment of Psychopathic Per-

sonality (CAPP). Int J Forensic Ment Health 11:242–252. https://doi.org/10.1080/14999013.2012.746759
Darke S, Kaye S, Finlay-Jones R, Hall W (1998) Factor structure of psychopathy among methadone maintenance patients. J Personal Disord 12:162–171. https://doi.org/10.1521/pedi.1998.12.2.162 (*)
Dotterer HL, Waller R, Neumann CS et al (2017) Examining the factor structure of the Self-Report of Psychopathy Short-Form across four young adult samples. Assessment 24:1062–1079. https://doi.org/10.1177/1073191116640355 (*)
Douglas KS, Vincent GM, Edens JF (2018) Risk of criminal recidivism: The role of psychopathy. In: Patrick CJ (ed) Handbook of Psychopathy, 2nd edn. Guilford, New York, pp 682–709
Eichenbaum AE, Marcus DK, French BF (2019) Item response theory analysis of the Psychopathic Personality Inventory–Revised. Assessment 26:1046–1058. https://doi.org/10.1177/1073191117715729 (*)
Eichenbaum AE, Marcus DK, French BF (2021) Item response theory analysis of the Triarchic Psychopathy Measure. Psychol Assess 33:766–776. https://doi.org/10.1037/pas0001022 (*)
Ellingson JM, Littlefield AK, Vergés A, Sher KJ (2018) Psychopathy and substance use disorder. In: Patrick CJ (ed) Handbook of psychopathy, 2nd edn. Guilford, New York, pp 635–661
Embretson SE, Reise SP (2000) Item response theory for psychologists. Lawrence Erlbaum, New Jersey
Fleiss JL (1986) Reliability of measurement. In: The design and analysis of clinical experiments. Wiley, New York, pp 1–32
Forth AE, Brown SL, Hart SD, Hare RD (1996) The assessment of psychopathy in male and female noncriminals: Reliability and validity. Pers Individ Dif 20:531–543. https://doi.org/10.1016/0191-8869(95)00221-9 (*)
George D, Mallery P (2003) SPSS for Windows step by step: A simple guide and reference/11.0 update, 4th edn. Allyn & Bacon, Boston
Gummelt HD, Anestis JC, Carbonell JL (2012) Examining the Levenson Self Report Psychopathy scale using a graded response model. Pers Individ Dif 53:1002–1006. https://doi.org/10.1016/j.paid.2012.07.014 (*)
Hanniball KB, Hohn RE, Fuller EK, Douglas KS (2021) Construct validity of the Comprehensive Assessment of Psychopathic Personality (CAPP): Examining the internal structure and generalizability of CAPP self-ratings across gender and ethnicity. Assessment 28:518–536. https://doi.org/10.1177/1073191120922621 (*)
Hare RD (1980) A research scale for the assessment of psychopathy in criminal populations. Pers Individ Dif 1:111–119. https://doi.org/10.1016/0191-8869(80)90028-8
Hare RD (1999) Psychopathy as a risk factor for violence. Psychiatr Q 70:181–197. https://doi.org/10.1023/A:1022094925150
Hare RD (2003) Hare Psychopathy Checklist–Revised (PCL-R): technical manual. Multi-Health Systems, Toronto
Hare RD, Neumann CS (2005) Structural models of psychopathy. Curr Psychiatry Rep 7:57–64. https://doi.org/10.1007/s11920-005-0026-3
Hare RD, Hart S, Cox DN (1995) The Hare Psychopathy Checklist: Screening Version (PCL-SV). Multi-Health Systems, Toronto
Hauck-Filho N, Teixeira MAP (2014) Revisiting the psychometric properties of the Levenson self-report psychopathy scale. J Pers Assess 96:459–464. https://doi.org/10.1080/00223891.2013.865196 (*)
Klein Haneveld E, Molenaar D, de Vogel V et al (2022) Do we hold males and females to the same standard? A measurement invariance study on the Psychopathy Checklist-Revised. J Pers Assess 104:368–379. https://doi.org/10.1080/00223891.2021.1947308 (*)
Leuschner F (2020) Täterinnen. Forens Psychiatr Psychol Kriminol 14:130–140. https://doi.org/10.1007/s11757-020-00590-4
Levenson MR, Kiehl KA, Fitzpatrick CM (1995) Assessing psychopathic attributes in a noninstitutionalized population. J Pers Soc Psychol 68:151–158. https://doi.org/10.1037/0022-3514.68.1.151
Lilienfeld SO (1990) Development and preliminary validation of a self-report measure of psychopathic personality. Unpublished doctoral dissertation, University of Minnesota, Minneapolis
Lilienfeld SO, Hess TH (2001) Psychopathic personality traits and somatization: Sex differences and the mediating role of negative emotionality. J Psychopathol Behav Assess 23:11–24. https://doi.org/10.1023/A:1011035306061
Lilienfeld SO, Widows MR (2005) Psychopathic Personality Inventory-Revised: professional manual. Psychological Assessment Resources, Lutz
Lynam DR, Whiteside S, Jones S (1999) Self-reported psychopathy: A validation study. J Pers Assess 73:110–132. https://doi.org/10.1207/S15327752JPA730108 (*)
Lynam DR, Gaughan ET, Miller JD et al (2011) Assessing the basic traits associated with psychopathy: Development and validation of the elemental psychopathy assessment. Psychol Assess 23:108–124. https://doi.org/10.1037/a0021146
Lynam DR, Sherman ED, Samuel D et al (2013) Development of a short form of the elemental psychopathy assessment. Assessment 20:659–669. https://doi.org/10.1177/1073191113502072
Meredith W (1993) Measurement invariance, factor analysis and factorial invariance. Psychometrika 58:525–543. https://doi.org/10.1007/BF02294825
Miller JD, Gaughan ET, Pryor LR (2008) The Levenson Self-Report Psychopathy Scale: an examination of the personality traits and disorders associated with the LSRP factors. Assessment 15:450–463. https://doi.org/10.1177/1073191108316888
Neal TMS, Sellbom M (2012) Examining the factor structure of the Hare Self-Report Psychopathy Scale. J Pers Assess 94:244–253. https://doi.org/10.1080/00223891.2011.648294 (*)
Neumann CS, Hare RD (2008) Psychopathic traits in a large community sample: Links to violence, alcohol use and intelligence. J Consult Clin Psychol 76:893–899. https://doi.org/10.1037/0022-006X.76.5.893 (*)
Neumann CS, Hare RD, Newman JP (2007) The super-ordinate nature of the Psychopathy Checklist-Revised. J Personal Disord 21:102–117. https://doi.org/10.1521/pedi.2007.21.2.102 (*)
Neumann CS, Schmitt DS, Carter R et al (2012) Psychopathic traits in females and males across the globe. Behav Sci Law 30:557–574. https://doi.org/10.1002/bsl.2038 (*)
Nicholls TL, Ogloff JR, Brink J, Spidel A (2005) Psychopathy in women: a review of its clinical usefulness for assessing risk for aggression and criminality. Behav Sci Law 23:779–802. https://doi.org/10.1002/bsl.678
Nye CD, Drasgow F (2011) Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. J Appl Psychol 96:966–980. https://doi.org/10.1037/a0022955
Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n7
Patrick CJ (2010) Operationalizing the triarchic conceptualization of psychopathy: preliminary description of brief scales for assessment of boldness, meanness, and disinhibition (Unpublished manual). Department of Psychology, Florida State University. https://pdfs.semanticscholar.org/fb4c/c0fdae0840553500ee091579ae76a6477e8e.pdf Accessed 15 Sep 2022
Paulhus DL, Neumann CS, Hare RD (2017) SRP-4: Self-Report Psychopathy Scale, 4th edn. Multi-Health Systems, Toronto
Pedersen L, Kunz C, Elsass P, Rasmussen K (2010) Psychopathy as a risk factor for violent recidivism: Investigating the Psychopathy Checklist Screening Version (PCL:SV) and the Comprehen-

sive Assessment of Psychopathic Personality (CAPP) in a forensic psychiatric setting. Int J Forensic Ment Health 9:308–315. https://doi.org/10.1080/14999013.2010.526681

Putnick DL, Bornstein MH (2016) Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev Rev 41:71–90. https://doi.org/10.1016/j.dr.2016.06.004

Reise SP, Ainsworth AT, Haviland MG (2005) Item Response Theory: Fundamentals, applications, and promise in psychological research. Curr Dir Psychol Sci 14:95–101. https://doi.org/10.1111/j.0963-7214.2005.00342.x

Salekin RT, Rogers R, Sewell KW (1997) Validity of psychopathy in a female offender sample: A multitrait-multimethod evaluation. J Abnorm Psychol 106:576–585. https://doi.org/10.1037//0021-843x.106.4.576

Salekin RT, Chen DR, Sellbom M et al (2014) Examining the factor structure and convergent and discriminant validity of the Levenson Self-Report Psychopathy Scale: Is the two-factor model the best fitting model? Personal Disord 5:289–304. https://doi.org/10.1037/per0000073

Sellbom M (2011) Elaborating on the construct validity of the Levenson self-report psychopathy scale in incarcerated and non-incarcerated samples. Law Hum Behav 35:440–451. https://doi.org/10.1007/s10979-010-9249-x (*)

Sellbom M, Cooke DJ (2020) Comprehensive Assessment of Psychopathic Personality–Self-Report (CAPP-SR), version 1: manual for administration, scoring and interpretation. https://capp-network.no/capp-sr/. Accessed 15 Sep 2022

Sellbom M, Cooke DJ, Hart SD (2015) Construct validity of the Comprehensive Assessment of Psychopathic Personality (CAPP) concept map: Getting closer to the core of psychopathy. Int J Forensic Ment Health 14:172–180. https://doi.org/10.1080/14999013.2015.1085112 (*)

Sellbom M, Lilienfeld SO, Fowler KA, McCrary KL (2018) The self-report assessment of psychopathy: Challenges, pitfalls, and promises. In: Patrick CJ (ed) Handbook of psychopathy, 2nd edn. Guilford, New York, pp 211–258

Skeem JL, Mulvey E, Grisso T (2003) Applicability of traditional and revised models of psychopathy to the Psychopathy Checklist: Screening Version. Psychol Assess 15:41–55. https://doi.org/10.1037/1040-3590.15.1.41 (*)

Somma A, Fossati A, Patrick CJ et al (2014) The three-factor structure of the Levenson Self-Report Psychopathy Scale: Fool's gold or true gold? A study in a sample of Italian adult non-clinical participants. Personal Ment Health 8:337–347. https://doi.org/10.1002/pmh.1267 (*)

Strand S, Belfrage H (2005) Gender differences in psychopathy in a Swedish offender sample. Behav Sci Law 23:837–850. https://doi.org/10.1002/bsl.674 (*)

Thomson ND, Bozgunov K, Psederska E, Vassileva J (2019) Sex differences on the four-facet model of psychopathy predict physical, verbal, and indirect aggression. Aggress Behav 45:265–274. https://doi.org/10.1002/ab.21816 (*)

Vandenberg RJ, Lance CE (2000) A review and synthesis of the measurement in-variance literature: Suggestions, practices, and recommendations for organizational research. Organ Res Methods 3:4–70. https://doi.org/10.1177/109442810031002

Verona E, Vitale J (2018) Psychopathy in women: Assessment, manifestations, and etiology. In: Patrick CJ (ed) Handbook of psychopathy, 2nd edn. Guilford, New York, pp 509–528

Walsh HC, Roy S, Lasslett HE, Neumann CS (2019) Differences and similarities in how psychopathic traits predict attachment insecurity in females and males. J Psychopathol Behav Assess 41:537–548. https://doi.org/10.1007/s10862-018-9704-4 (*)

Walters GD, Wilson NJ, Glover AJ (2011) Predicting recidivism with the Psychopathy Checklist: Are factor score composites really necessary? Psychol Assess 23:552–557. https://doi.org/10.1037/a0022483 (*)

Williams KM, Paulhus DL, Hare RD (2007) Capturing the four-factor structure of psychopathy in college students via self-report. J Pers Assess 88:205–219. https://doi.org/10.1080/00223890701268074

Windle M, Dumenci L (1999) The factorial structure and construct validity of the Psychopathy Checklist-Revised (PCL-R) among alcoholic inpatients. Struct Equ Modeling 6:372–393. https://doi.org/10.1080/10705519909540141 (*)

Yoon D, Mokros A, Rettenberger M et al (2022) Triarchic psychopathy measure: convergent and discriminant validity in a correctional treatment setting. Pers Disord Theory Res Treat 13:52–63. https://doi.org/10.1037/per0000478